

ISSN 2088-9984

seminar nasional & ekspo teknik elektro 2012

Widrasuti

Medan, 14 November 2012

Organizer :



JURUSAN TEKNIK ELEKTRO
UNIVERSITAS SYIAH KUALA



DEPARTEMEN TEKNIK ELEKTRO
UNIVERSITAS SUMATERA UTARA

PROSIDING

Sponsor :



PT. SUKSES PRIMA RAYA



Co-organizer :

Wid

PROSIDING
SNETE 2012

SEMINAR NASIONAL DAN EKSPLO TEKNIK ELEKTRO 2012

ISSN: 2088-9984

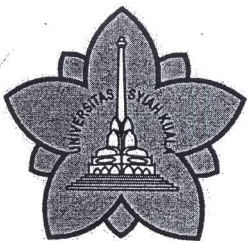
14 NOVEMBER 2012

SANTIKA DYANDRA HOTEL & CONVENTION

MEDAN

Editor:

Dr. Ir. Rizal Munadi M.M.,M.T.



**JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS SYIAH KUALA**



**DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS SUMATERA UTARA**

KATA PENGANTAR

Mewakili panitia Seminar Nasional dan Ekspo Teknik Elektro (SNETE), saya menyampaikan salam hangat kepada seluruh peserta seminar ini, khususnya peserta SNETE 2012 yang telah menyumbangkan pikiran dan karyanya dalam kegiatan ini. Kegiatan seminar ilmiah ini merupakan kelanjutan dari tahun sebelumnya yang diadakan di Banda Aceh. Ketika itu, Jurusan Teknik Elektro Unsyiah bekerjasama dengan Prodi Teknik Elektronika Politeknik Aceh sebagai penyelenggara bersama. Pada tahun 2012 yang merupakan ajang kedua kalinya SNETE dilaksanakan dan terealisasi berkat kerjasama Jurusan Teknik Elektro Universitas Syiah Kuala bersama Departmen Teknik Elektro Universitas Sumatera Utara. Tema SNETE tahun ini adalah *Penguatan Peran Pendidikan Tinggi Teknik Elektro dalam Mengisi Kebutuhan Pembangunan dan Industri*.

Alhamdulillah, keinginan panitia melibatkan pihak dunia industri mendapat sambutan yang hangat. Pada kesempatan tahun ini, hadir beberapa pembicara kunci dari kalangan swasta di bidang telekomunikasi dan energi. Mudah-mudahan kehadiran pembicara kunci ini dapat memberikan manfaat bagi dunia pendidikan tinggi Teknik Elektro.

Sebagai salah satu cabang keilmuan eksakta, teknologi di bidang Teknik Elektro mengalami perkembangan yang cukup pesat. Teknologi telepon seluler misalnya, telah mampu mengalahkan penetrasi jumlah pemakai telepon berbasis sambungan tetap atau kabel. Konvergensi teknologi telekomunikasi dan komputer telah melahirkan dunia informasi yang lebih efisien. Salah satunya teknologi internet. Demikian pula perkembangan bidang lainnya di bidang teknik elektro seperti elektronika, kontrol dan energi listrik. Semuanya menjadi sangat penting dan berarti dalam kehidupan modern saat ini. Oleh karenanya, inovasi dan karya di bidang ini perlu dipelajari, dikembangkan, serta dapat diimplementasikan. Untuk itulah, forum kegiatan ilmiah dalam bentuk seminar ini menjadi penting untuk diikuti sebagai sarana tukar informasi dan kepakaran diantara sesama peneliti dan kalangan dunia industri.

Atas nama panitia SNETE 2012, izinkan saya menyampaikan penghargaan atas dukungan terselenggaranya kegiatan ini kepada: Rektor Universitas Syiah Kuala, Dekan Fakultas Teknik Universitas Syiah Kuala, Rektor Universitas Sumatera Utara, Dekan Fakultas Teknik Universitas Sumatera Utara, serta institusi pendidikan lainnya yang mendukung acara ini: Universitas Malikussaleh, Politeknik Negeri Lhokseumawe, Universitas Al-Muslim, Politeknik Aceh; para pembicara kunci, para sponsor, dan seluruh panitia pelaksana sehingga seminar ini dapat terlaksana dengan baik. Secara khusus, saya juga mengucapkan terima kasih kepada seluruh pemakalah dan peserta, atas partisipasi, kontribusi dan inovasi nyata bagi perkembangan ilmu Teknik Elektro melalui SNETE2012. Selamat Seminar!

Medan, 14 November 2012

Dr. Ir. Rizal Munadi, MM., MT.
Ketua Panitia

PANITIA SEMINAR NASIONAL DAN EKSPLO TEKNIK ELEKTRO

SNETE 2012

Penanggung Jawab	: 1. Dr. Ir. Marwan (Dekan Fakultas Teknik Universitas Syiah Kuala) 2. Prof.Dr.Ir. Bustami Syam, M.S.M.E (Dekan Fakultas Teknik Universitas Sumatera Utara)
Wakil Penanggung Jawab	: 1. Dr. Ir. Mirza Irwansyah, MBA. MLA (Pembantu Dekan I, Fakultas Teknik - Unsyiah) 2. Ir. Syahrizal, MT (Pembantu Dekan II, Fakultas Teknik - Unsyiah) 3. Dr. M. Ilham Maulana, ST., MT (Pembantu Dekan III, Fakultas Teknik - Unsyiah) 4. Dr. Ir. Taufiq Saidi, M.Eng (Pembantu Dekan IV, Fakultas Teknik - Unsyiah)
Ketua Pelaksana	: Dr. Ir. Rizal Munadi, MM., M.T. (Ketua Jurusan Teknik Elektro – Universitas Syiah Kuala)
Wakil Ketua Pelaksana	: Ir. Surya Tarmizi Kasim, M.Si. (Ketua Departemen Teknik Elektro – Universitas Sumatera Utara)
Sekretaris	: Sayed Muchallil, ST., M. Sc. (Unsyiah)
Bendahara	: M. Irhamsyah, ST., MT. (Unsyiah)
Komite Pelaksana	:

Program

1. Tarmizi, S.T., M.Sc. (Unsyiah)
2. Rahmad Fauzi, S.T., M.T. (USU)
3. Zulfikar, S.T., M.Sc. (Unsyiah)
4. Syukriyadin, S.T., M.T. (Unsyiah)
5. Dewi Yana, S.Hi. (Unsyiah)

Publikasi dan Dokumentasi

1. Yudha Nurdin, S.T., M.T. (Unsyiah)
2. Ir. Syahrawardi (USU)
3. Hubbul Walidainy, S.T., M.T. (Unsyiah)
4. Fardian, S.T., M.Sc. (Unsyiah)

Logistik

1. Mahdi Syukri, S.T., M.T. (Unsyiah)
2. Ir. Zulkarnaen Pane (USU)
3. Alfisyahrin , S.T., M.T. (Unsyiah)
4. Melinda, S.T., M.Sc. (Unsyiah)
5. Yudha Iskandar, S.T. (Unsyiah)
6. Suhendra (Unsyiah)

Kesekretariatan

1. Yunidar, S.Si., M.T. (Unsyiah)
2. Syiska Yana, S.T., M.T. (USU)
3. Ir. Walidin, M.Sc. (Unsyiah)
4. Roslidar, S.T., M.S.Tc.E. (Unsyiah)
5. Alfatirta Mufti, S.T., M.Sc. (Unsyiah)
6. Zul Syukri, S.T. (Unsyiah)
7. Jasmiati, A.Md. (Unsyiah)

Expo

1. Ramdhan Halid Siregar, S.T., M.T. (Unsyiah)
2. Ali Hanafiah Rambe, S.T., M.T. (USU)
3. Ahmadiar, S.T., M.Sc. (Unsyiah)
4. Ali Imron, S.T. (Unsyiah)
5. Edi Sukriansyah, S.T. (Unsyiah)

DAFTAR ISI

	Hal.
Kata Pengantar	ii
Panitia Seminar Nasional dan Expo Teknik Elektro SNETE 2012	iii
Daftar Isi	iv
A. ENERGI LISTRIK	
Analisis Aliran Daya Beban Tidak Seimbang Pada Feeder Blang Bintang GH Lambaro Banda Aceh <i>Syahrizal, Syukriyadin, Ridha Firdaus</i> <i>Universitas Syiah Kuala</i>	A-1
Studi Pemodelan Integrasi Pembangkit Skala Mikro Terdistribusi pada Daerah Isolated di Aceh <i>Fajrul Ghafur Nst, Syukriyadin</i> <i>Universitas Syiah Kuala</i>	A-8
Identifikasi Gangguan Kualitas Daya Pada Sistem Tenaga Listrik Menggunakan FFT dan Neural Network <i>I Gede Dyana Arjana, I Nyoman Budiastra, I Nyoman Setiawan</i> <i>Universitas Udayana</i>	A-14
Strategi Perbaikan Tegangan Injeksi Dvr Menggunakan Proporsional Integral Posicast Kontroller Pada Sistem Distribusi 20 kV <i>Ezwarsyah, Asri</i> <i>Universitas Malikussaleh</i>	A-20
Pengaruh Kerja Recloser Pada Distribusi Tegangan Transien Rumah Tangga <i>Andik Bintoro</i> <i>Universitas Malikussaleh</i>	A-26
Implementasi Sistem Smartphone untuk Penghematan Energi Listrik <i>I Nyoman Budiastra, I Gde Dyana Arjana, I Nyoman Setiawan</i> <i>Universitas Udayana</i>	A-30
Pengendalian Arus Starting Air Conditioning (AC) Berbasis Mikrokontroler ATMEGA8535 <i>I Wayan Arta Wijaya, Cok. Gede Indra Partha</i> <i>Universitas Udayana</i>	A-34
Penggunaan Algoritma Floating Point untuk Optimasi Rugi-Rugi Pada Parameter Motor Induksi <i>Birowo, Darjat S, Ishak, Marsono</i> <i>STIMIK Triguna Dharma Medan</i>	A-39

- Studi Perbaikan Faktor Daya Beban Induktif dengan Kompensator Reaktif Seri Menggunakan Sakelar Pemulih Energi Magnetik A-42
Fauzan
Politeknik Negeri Lhokseumawe
- Optimization of Hybrid PV/Wind/Diesel System for Remote Rural Electrification A-48
Suriadi, Soib Taib, Mohd Shawal Jadin
Universitas Syiah Kuala
- Advanced Infrared Image Processing Method for Evaluating the Thermal Condition of Electrical Installation A-53
Mohd Shawal Jadin, Soib Taib, Suriadi
Universiti Malaysia Pahang
- Pemanfaatan Urine Manusia Sebagai Energi Alternatif Bahan Bakar Minyak A-60
Donny Ajie Baskoro, Faisal Yafi, Denmas Muhammad Ridwan, Ardhani Reswari Yudistari, Fuji Ihsani
Universitas Gunadarma
- Rancang Bangun Pembangkit Listrik Tenaga Pikohidro Sistem Terapung A-63
Mahdi Syukri, Ramdhan Halid, Harry Sukma
Universitas Syiah Kuala
- Simulasi Integrasi PLTB di Daerah Pidie Jaya dengan Sistem Grid 220V PLN A-70 Menggunakan PSCAD/EMTDC
Edward Ammi Djamin, Syukriyadin, Hamdani
Universitas Syiah Kuala
- Pengukuran Unjuk Kerja Sistem Pompa Air-Energi Surya A-76
Hamdani, Irwansyah, Ilyas
Universitas Syiah Kuala
- Evaluasi Voltage Sag Sistem Tenaga Listrik Akibat Gangguan Hubung Singkat A-80
Syafii, Novi Syafitri
Universitas Andalas
- Pengendalian Arus Starting Air Conditioning (AC) Berbasis Mikrokontroler A-86 ATMEGA8535
I Wayan Arta Wijaya, Cok. Gede Indra Partha
Universitas Udayana
- Sistem Kontrol Pemakaian Energi Listrik dengan Skala Prioritas Menggunakan Mikrokontroler ATMEGA8535 A-91
Cok. Gede Indra Partha, I Wayan Arta Wijaya
Universitas Udayana
- Analisis Aliran Daya Sistem Tenaga Listrik Berbasis Geographic Information System A-96
Syukriyadin, Mansur Gapy, Walidin
Universitas Syiah Kuala

B. TEKNIK TELEKOMUNIKASI

- Characterization of Microstrip Circular Patch Antenna with Dielectric Resonator B-1
Achmad Munir, Antrisha Daneraici Setiawan, Mohammad Sigit Arifianto
Institut Teknologi Bandung
- Pemodelan Matematis Struktur Planar Waveguide dengan Transmisi Line Method (TLM) B-5
Muhammad Ikhwanus
Universitas Malikussaleh
- Kajian Sistem Monitoring dan Informasi Peringatan Tsunami Menggunakan Wireless UHF Camera dan Wireless IP Camera B-11
Edi Sukriansyah, Hubbul Walidainy, Nasaruddin
Universitas Syiah Kuala
- Evaluasi Hotspot Gratis di Kota Banda Aceh Menggunakan Netstumbler B-16
Rizal Munadi, Kesuma Fitri, Ernita Dewi Meutia
Universitas Syiah Kuala

C. ELEKTRONIKA

- Sistem Penampil Informasi Parkir Kendaraan Roda Empat Berbasis Mikrokontroller AT89S51 C-1
Alfisyahrin, Fajri
Universitas Syiah Kuala
- Perancangan dan Pembuatan Kandang Otomatis Untuk Kura-Kura Darat C-9
Marvin Chandra Wijaya, Felix, Semuil Tjihardi
Universitas Kristen Maranatha
- Pemodelan Polisi Tidur Otomatis C-15
Andrew Sebastian Lehman, Semuil Tjiharjadi, Oscar Halim Gozali
Universitas Kristen Maranatha
- Desain dan Layout Komparator Presisi pada ADC Pipeline 1-bit/stage untuk Aplikasi Kamera Kecepatan Tinggi C-19
Hamzah Afandi, Erma Triawati Ch, Atit Pertiwi
Universitas Gunadarma
- Sistem Kontrol Parkir Mobil Otomatis Terkomputerisasi dan Berbasis Mikrokontroler ATMEGA 16 C-25
Cok. Gede Indra Partha, I B Alit Swamardika
Universitas Udayana
- Pengukuran Asap Buang Hasil Pembakaran Sampah Organik & Non Organik untuk Menentukan Peringatan Dini Kebakaran Hutan C-31
Arif Gunawan
Politeknik Caltex Riau

Concentration and Velocity Measurement of Two Phase Flow Using Optical and Ultrasonic Tomography Sallehuddin Ibrahim, Mohd. Sazli Saad, Mohd. Amri Md Yunus <i>Universiti Teknologi Malaysia</i>	C-37
Rancang Bangun Gasing Khas Yogyakarta (Studi Kasus : Gasing Adu Bunyi) Satria Giri Nugraha, Agus Cahyo Nugroho, Suyoto <i>Universitas Atma Jaya Yogyakarta</i>	C-42
Kendali Fuzzy Stimulasi Elektrik Penyembuhan Luka Berdasarkan Penilaian PUSH TotalScore Rachmawati, Achmad Arifin, Gunawan <i>Politeknik Negeri Lhokseumawe</i>	C-47
Desain Inverter 700W Berbasis Mikroprosesor untuk Sistem Photovoltaic (PV) Mandiri Wan Zain, Soib Taib, Suriadi, M. Nazir <i>Universitas Sains Malaysia</i>	C-51
Desain Bank Kapasitor Optimal Berbasis Assesmen Stokastik Menggunakan Particle Swarm Optimization I Ketut Suryawan <i>Politeknik Negeri Bali</i>	C-57
Sistem Pengaman Kebakaran Pada Rumah Cerdas Dengan Detektor Asap Dan Panas Berbasis PLC OMRON CPM2A Agus Adria, Baiquny, Yunidar <i>Universitas Syiah Kuala</i>	C-65

D. TEKNIK INFORMATIKA DAN KOMPUTER

Perancangan Sistem Penunjang Keputusan Pemilihan Rumah Ramah Lingkungan Berbasis GBCI Surya Thiono Wijaya, Tryono Taqwa, Yusuf Triyuswoyo <i>Universitas Gunadarma</i>	D-1
Perencanaan Routing dengan Permintaan Acak Menggunakan Chance Constrained Programming Sardes Malau, Tulus <i>Universitas Sumatera Utara</i>	D-8
Peranan Sistem Informasi Berbasis Web untuk Manajemen Data Meter Elektronik PT PLN (Persero) Cabang Barabai Guson Prasamuarso Kuntarto <i>Universitas Bakrie</i>	D-13
Fire Alarm Management Information System Yusuf Lestanto <i>Universitas Bakrie</i>	D-20

Analisis Serangan dengan <i>Selective Plaintext</i> pada Sebuah Algoritma Enkripsi Citra Berbasis <i>Chaos</i> Rinaldi Munir <i>Institut Teknologi Bandung</i>	D-24
Pengembangan Aplikasi Distribusi Internet Berbasis Wireless Fidelity Dengan Menggunakan Bahasa Pemrograman PHP Muhammad Safri Lubis, Erlin Umar Dani <i>Universitas Sumatera Utara</i>	D-29
Pengenalan Huruf Tulisan Tangan Menggunakan Metode Clustering Berdasarkan Bentuk Segmen Lulu C. Munggaran, Suryarini Widodo, Nuryuliani, Riski Purwo Handariningsih <i>Universitas Gunadarma</i>	D-36
Analisis Kapasitas dan Imperceptibility Proses Penyisipan Pesan Pada Citra Menggunakan LWT dan Metode BPCS Malahayati, Benny B. Nasution, Fitri Arnia <i>Universitas Syiah Kuala</i>	D-43
Pemanfaatan Rich Site Summary (RSS) Untuk Aplikasi Portal Berita Pada Telepon Selular Menggunakan Java ME (Micro Edition) Ihsan Jatnika, Fivi Syukriah, Fitria Rohmana Rogawa, Irmira Setyaningrum <i>Universitas Gunadarma</i>	D-48
Penggunaan Metode Support Vector Machine untuk Mengklasifikasi dan Memprediksi Angkutan Udara Jenis Penerbangan Domestik dan Penerbangan Internasional di Banda Aceh Sayed Fachrurrazi <i>Universitas Malikussaleh</i>	D-55
Perbandingan Performansi Asynchronous dan Synchronous Message Passing Pada Komputasi Paralel Himmat Rija, Taufiq Abdul Gani, Melinda, Yuwaldi Away <i>Universitas Syiah Kuala</i>	D-59
Pembuatan Sistem Aplikasi Komunikasi Data Multi-Tier dengan Konsep Agent Nikolaus Very Permana, Berkah I. Santoso, Widia Nursiyanto <i>Universitas Multimedia Nusantara</i>	D-65
Model Konseptual Aplikasi SIN (Single Identity Number) untuk ARC (Aceh Reference Code) Melinda, Mirza Rizaldy, T. Robby Irza, Yuwaldi Away <i>Universitas Syiah Kuala</i>	D-71
Penerapan Teknologi Radio Frequency Identification (RFID) untuk Pengendalian Kinerja Karyawan Muhammad Aiyub, Yuwaldi Away, Melinda <i>Universitas Syiah Kuala</i>	D-78

Pendeteksian dan Penangkapan Gerakan Fitur Wajah Secara 3D Menggunakan Active Appearance Models (AAM) Atthariq <i>Politeknik Negeri Lhokseumawe</i>	D-82
Pengembangan Aplikasi Mobile Pembuatan Ketupat dan Perhitungan Kalorinya Berbasis Android Suyoto, Nyoman Ayu Nila Dewi, Ragil Tri Dianti Putri <i>Universitas Atmajaya Yogyakarta</i>	D-88
Teknologi Scanner dan Security System untuk Mengurangi Antrian Pembelian Barang Alif Ahmad Syamsudduha, Muhammad Rendianto, Rizka Khairunnisa, Lily Wulandari <i>Universitas Gunadarma</i>	D-93
Perancangan Sistem Pemilihan Rute Alternatif dengan Penyesuaian Waktu, Situasi dan Kondisi Muhammad Rendianto, Nurdianah Fitri, Rizka Khairunnisa, Lily Wulandari <i>Universitas Gunadarma</i>	D-99
Pemodelan dan Simulasi Evakuasi Bencana Tsunami Berbasis Multi Agent System Rizqiya Windy Saputra, Khairul Munadi, Yudha Nurdin <i>Universitas Syiah Kuala</i>	D-105
Sistem Ujian Teori Online Laboratorium Komputer Menggunakan Framework CodeIgniter Erlina, Revida Iriana Napitupulu, Lulu C. Munggaran, Kartika Kurniarin <i>Universitas Gunadarma</i>	D-111
Perhitungan Luas Area Hutan pada Citra dengan Algoritma Segmentasi Warna Lokal Dyah Pratiwi, Karmilasari, Sarifuddin M., Lussiana E.T.P. <i>Universitas Gunadarma</i>	D-116
Perancangan Sistem Pemilu Touch Screen dengan Pengiriman Suara Real Time Dennis Aprilla Christine, Jonathan Hindharta, Yoga Perdana Sasmita, Virgiawan Ananda Pratama, Lily Wulandari <i>Universitas Gunadarma</i>	D-121
Barcode Shopping Cart Faisal Yafi, Michael Satrio, Mohammad Ropiyudin, Nurdianah Fitri, Lily Wulandari <i>Universitas Gunadarma</i>	D-126
Perancangan Sistem Augmented Reality Recording Moment (ARReMo) Denmas Muhammad Ridwan, Istiana Idha Aulia, Odheta, Sandi Agung Harseno, Lily Wulandari <i>Universitas Gunadarma</i>	D-130
Analisis Metode Alokasi Fragmen Pada Sistem Basis Data Terdistribusi Miftahul Jannah, Baby Lolita, Hustinawati <i>Universitas Gunadarma</i>	D-134
Perancangan Sistem Informasi Alat Travel Guide "Walk Around The World" Ario Halik, Septiawan, Soeltan Zaki Rizaldy, Lily Wulandari <i>Universitas Gunadarma</i>	D-140

- Perancangan Sistem Informasi Video Conference untuk Mendukung Rapat
Ferina Ferdianti, Lia Ambarwati, Melisa Chatrine Kamu, Paramitha Megarani, Lily Wulandari
Universitas Gunadarma D-144
- Perancangan Sistem Informasi Praktikum Komputer Berbasis Client Server
Widiastuti, Kemal Ade Sekarwati
Universitas Gunadarma D-150
- Perbandingan Natural Language Processing untuk Pra Pemrosesan Teks Bahasa Inggris
Kemal Ade Sekarwati, Widiastuti
Universitas Gunadarma D-157

Perbandingan Natural Language Processing Untuk Pra Pemrosesan Teks Bahasa Inggris

Kemal Ade Sekarwati ¹⁾ Ana Kurniawati ²⁾

^{1,2} Jurusan Sistem Informasi, Fakultas Ilmu Komputer & Teknologi Informasi, Universitas Gunadarma
Jl. Margonda Raya 100 Depok – Jawa Barat 16424

¹ade@staff.gunadarma.ac.id, ²ana@staff.gunadarma.ac.id

ABSTRAK

Kemudahan berbagi informasi melalui internet telah mendorong pencarian literatur secara online. Hal ini menyebabkan banyak orang terutama di bidang akademik meniru gagasan orang lain atau bekerja tanpa menyebutkan sumbernya. Oleh karena itu pencegahan masalah seperti ini merupakan hal yang sangat penting untuk dilakukan terutama untuk tujuan pendidikan. Pendekatan deteksi plagiarisme sebuah dokumen yang dicurigai dengan dokumen asli dapat dilakukan dengan menggunakan teknik natural language processing (NLP). NLP melibatkan pengolahan bahasa manusia oleh mesin. Berbagai bidang komputer seperti bantuan pelajaran bahasa dan ekstraksi informasi biomedis sudah memberikan pengalaman yang bermanfaat dari penggunaan NLP. NLP merupakan area yang masih kurang dieksplorasi untuk pendeteksian plagiarisme. Pada paper ini akan dibahas hasil analisis dari pendekatan teknik natural language processing untuk melakukan pra pemrosesan teks pada dokumen berbahasa Inggris.

Kata Kunci: deteksi plagiarisme, natural language procesing, pra pemrosesan, bahasa Inggris.

1. Pendahuluan

Masalah plagiarisme telah ada dalam waktu yang lama tetapi dengan kemajuan teknologi informasi masalahnya menjadi lebih buruk. Hal ini karena terdapat banyak materi versi elektronik tersedia bagi siapapun. Web merupakan sumber yang penting dan umum untuk plagiarisme. Beberapa program deteksi plagiarisme seperti Turnitin dikembangkan untuk mencoba untuk menangani masalah ini. Untuk menentukan apakah sebuah artikel disalin dari sumber-sumber elektronik web atau lainnya, deteksi plagiarisme harus menghitung kemiripan antara dua artikel. Namun, seringkali sulit untuk mendeteksi plagiarisme secara akurat setelah modifikasi isi disalin. Sebagai contoh, dimungkinkan untuk cukup mengganti sebuah kata dengan sinonimnya (misalnya "Program"

dengan "Perangkat lunak") dan mengubah seluruh struktur kalimat.

Kemudahan berbagi informasi melalui internet telah mendorong mencari literatur *online* yang menyebabkan banyak orang, terutama di bidang akademik meniru gagasan orang lain atau bekerja tanpa menyebutkan sumbernya. Oleh karena itu pencegahan masalah seperti ini merupakan hal yang sangat penting untuk dilakukan terutama untuk tujuan pendidikan. Selama bertahun-tahun banyak metodologi telah dikembangkan untuk melakukan deteksi otomatis plagiarisme, termasuk alat-alat untuk deteksi teks bahasa alami seperti Turnitin, CopyCatch dan alat untuk mendeteksi sumber kode komputer pemrograman seperti MOSS (*Measure Of Software Similarity*) yang diperkenalkan oleh Aiken pada tahun 1994 [1].

Deteksi plagiarisme bukan area penelitian baru. Berbagai pendekatan telah dikembangkan untuk menangani dengan baik plagiarisme eksternal dan hakiki/intrinsik pada teks-teks tertulis [8]. Deteksi plagiarisme eksternal berisi perbandingan dokumen plagiarisme yang mencurigakan terhadap dokumen asli. Di sisi lain, deteksi plagiarisme intrinsik, berisi penemuan bagian-bagian dalam dokumen plagiarisme tanpa akses ke teks asli.

Metode yang digunakan untuk deteksi plagiarisme sejauh ini sebagian besar terbatas pada tingkat yang sangat dangkal, misalnya dengan membandingkan teks-teks yang mencurigakan dan teks asli di tingkat string untuk memeriksa jumlah kata yang tumpang tindih antar dokumen [3]. Meskipun karya terbaru [4] telah menargetkan teknik pra pemrosesan untuk menggeneralisasi dokumen dengan menggantikan kata-kata dalam bentuk dasar, teknik yang digunakan masih sangat terbatas dan tidak ada perbaikan yang signifikan yang dilaporkan. Oleh karena itu plagiarisme terus menjadi tantangan yang berkembang, yang mempengaruhi banyak bidang diantaranya bidang pendidikan, penerbitan dan bahkan sektor usaha.

Pada makalah ini akan dibahas hasil analisis dari pendekatan teknik *natural language processing* (nlp) untuk melakukan pra pemrosesan teks pada dokumen berbahasa

Inggris. Pendekatan teknik nlp untuk pra pemrosesan dokumen bahasa Inggris yang akan dibahas yaitu pendekatan yang dilakukan oleh Ceska dan Fox [4] serta Runeson, Alexandersson dan Nyholm [9].

2. Pendekatan NLP Untuk Mendeteksi Plagiarisme

Penelitian yang dilakukan oleh Clough [6] dan Androutsopoulos dan Malakasiotis [2] menunjukkan hal yang diperlukan untuk menerapkan teknik NLP pada plagiarisme dan hal ini dapat menghasilkan akurasi yang lebih baik melalui deteksi naskah yang beberapa katanya telah mengalami perubahan. Namun, tidak ada percobaan yang berhasil dilakukan untuk menunjukkan bahwa hal ini memang terjadi.

Ceska dan Fox [4] menerapkan beberapa teknik pra-pemrosesan untuk meningkatkan akurasi deteksi plagiarisme. Teknik ini termasuk *heuristic* yang sederhana seperti mengganti angka dengan simbol *dummy*, menghilangkan tanda baca, penerapan teknik-teknik NLP dasar seperti *lemmatisation*, penghapusan kata-kata yang tidak relevan dan penggabungan kamus untuk menggeneralisasi kata-kata pada naskah.

Runeson, Alexandersson dan Nyholm [9] menggunakan pra-pemrosesan NLP dengan naskah pendek untuk mendeteksi laporan yang digandakan, termasuk *tokenisation*, akar kata, perhentian-penghapusan kata.

Chong, Specia dan Mitkov [5] menggunakan kerangka kerja untuk deteksi plagiarisme eksternal di mana sejumlah teknik NLP diterapkan untuk memproses satu set dokumen yang mencurigakan dan asli, tidak hanya untuk menganalisis string tetapi juga struktur teks, menggunakan sumber daya untuk menjelaskan hubungan teks. Hasil awal yang diperoleh dengan korpus paragraf pendek yang ditiru menunjukkan bahwa NLP teknik meningkatkan akurasi pendekatan yang ada. Pendekatan yang dilakukan ini masih belum mendeteksi plagiarisme untuk multibahasa, belum dapat melakukan generalisasi sinonim (disambiguasi makna kata) dan generalisasi struktur kalimat.

Leung dan Yuen [7] menggunakan struktur sintaksis dan arti semantik dari dua kalimat yang dapat dibandingkan untuk menyatakan kemiripan kalimat-kalimat tersebut. Ada beberapa langkah penyesuaian pada prosedur. Pertama, tesaurus (atau struktur hirarkis leksikal) direferensikan untuk mengetahui sinonim, istilah yang lebih luas dan sempit yang digunakan dalam paper diperiksa. Kemudian, paper tsb akan dibandingkan dengan dokumen-dokumen dalam basis data. WordNet merupakan tipikal contoh tesaurus yang dapat digunakan untuk tujuan ini. Jika diduga bahwa paper memuat beberapa isi dari basis data, kalimat-kalimat dari paper dapat diuraikan untuk membangun pohon *parsing* dan representasi semantik untuk selanjutnya dibandingkan secara detail. Konteks tata bahasa bebas dan kasus tata bahasa digunakan

untuk merepresentasikan struktur sintaksis dan arti semantik dari kalimat dalam sistem. Hal ini ditemukan bahwa plagiarisme yang tidak dapat dideteksi dengan metode tradisional bisa diidentifikasi dengan pendekatan baru ini. Meskipun metode sintaksis dan semantik dapat membantu dalam memahami makna kalimat, waktu yang dibutuhkan untuk memproses kalimat menjadi lebih lama. Dengan demikian, disarankan bahwa pendekatan ini dapat diterapkan untuk mendeteksi plagiarisme ketika dokumen sumber sangat besar kemungkinannya untuk disalin. Misalnya, tugas ujian di masa yang lalu di sebuah universitas mungkin akan disalin oleh mahasiswa di universitas yang sama.

Tabel 1. Pendekatan NLP Untuk Mendeteksi Plagiarisme

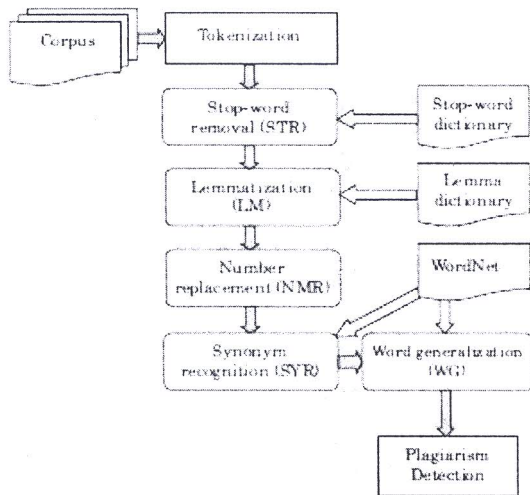
No	Peneliti, Tahun	Metode
1	Ceska & Fox (2009)	teknik pra-pemrosesan untuk meningkatkan akurasi deteksi plagiarisme
2	Runeson, Alexandersson & Nyholm (2007)	pra-pemrosesan NLP dengan naskah pendek untuk mendeteksi laporan yang digandakan, tokenisasi, <i>stemming</i> , dan <i>stop-word removal</i>
3	Miranda Chong, Lucia Specia, & Ruslan Mitkov (2010)	Menerapkan sejumlah teknik NLP untuk menganalisis string, struktur teks, dan menjelaskan hubungan teks
4	Chi Hong Leung & Yuen Yan Chan (2007)	Menggunakan metode pemrosesan sintaksis dan semantik untuk mengidentifikasi isi salinan yang telah dirubah struktur kalimatnya dan mengalami penggantian kata

2.1 Pra Pemrosesan

Pra pemrosesan teks mempunyai pengaruh yang signifikan terhadap kinerja beberapa tugas NLP termasuk pendeteksian plagiarisme. Deteksi plagiarisme merupakan area berbeda yang harus diberikan perhatian khusus yang tepat untuk diterapkan pada teknik pemrosesan yang luas jangkauannya. Berbagai pra pemrosesan mempunyai pengaruh yang berbeda-beda, seperti meningkatkan keakuratan, mempercepat waktu yang dibutuhkan, dan keduanya yaitu meningkatkan keakuratan dan mempercepat waktu yang dibutuhkan. Deteksi plagiarisme dapat menggunakan berbagai macam teknik pra pemrosesan dengan tujuan untuk meningkatkan keakuratan atau mengurangi langkah-langkah yang harus diproses.

2.2 Pendekatan Ceska dan Fox

Ceska dan Fox [4] memaparkan teknik pra pemrosesan seperti yang ditunjukkan pada Gambar 1. Blok yang paling penting adalah tokenisasi. Blok lainnya menggambarkan teknik lain yang dapat diterapkan jika pengguna menginginkannya.



Gambar 1. Skema Pra Pemrosesan Teks

Berikut ini penjelasan dari skema pra pemrosesan teks seperti pada Gambar 1 :

1. Tokenization

Proses *tokenization* akan melakukan ekstraksi kata tunggal dari teks berstruktur. Tanda baca juga dapat diekstraksi jika tanda baca tersebut dibutuhkan oleh proses lain.

2. Stop-word removal (STR)

STR merupakan pendekatan dasar pra pemrosesan yang melakukan penghapusan kata-kata umum. Untuk deteksi plagiarisme dapat menjadi rumit apabila menghapus kata-kata umum yang biasa pengarang gunakan, sehingga pengaruh STR tidak dapat diprediksi. Cara umum untuk membatasi apa saja yang dihitung sebagai STR adalah dengan menggunakan sebuah kamus yang berisikan daftar kata yang umum.

3. Lemmatization (LM)

Lemmatization merupakan proses menentukan bentuk dasar dari sebuah kata yang diketahui. Selama proses ini, hubungan kata digunakan untuk menentukan arti kata.

4. Number Replacement (NMR)

Number replacement merupakan pendekatan khusus yang mengganti semua angka dengan simbol *dummy*.

5. Synonym Recognition (SYR)

Jika sejumlah kata digantikan dengan sinonim maka metode deteksi akan gagal. *WordNet* digunakan sebagai kamus yang menyimpan kata-kata yang mempunyai arti yang sama.

6. Word Generalization (WG)

Word generalization merupakan proses penempatan bermacam-macam kata khusus dengan kata yang lebih umum. Sebagai contoh : kata "dog" dan "cat" digantikan dengan kata "animal".

2.3 Pendekatan Runeson, Alexandersson & Nyholm

Pendekatan pra pemrosesan yang dilakukan oleh Runeson hampir sama dengan pendekatan yang dilakukan Ceska dan Fox dengan penambahan langkah yaitu *vector space representation* dan *similarity calculation*. Langkah-langkah pra pemrosesan yang dilakukan oleh Runeson meliputi :

1. Tokenization

Seluruh tanda baca diabaikan. Kata penghubung tanpa spasi dipisahkan ke dalam beberapa token. Seluruh huruf besar dirubah menjadi huruf kecil.

2. Stemming

Tujuan dari *stemming* adalah mengidentifikasi bentuk dasar dari setiap kata. Kata-kata dapat ditulis dalam bentuk tata bahasa yang berbeda tetapi masih bisa menggambarkan informasi yang sama. Selama fase ini, imbuhan dan komponen leksikal dihapus dari setiap token dan hanya akar kata yang tersisa.

3. Stop words removal

Berisikan kata yang secara umum digunakan ditambah kata-kata yang sering muncul tapi tidak mempunyai arti penting.

4. Vector space representation/model

Vector space model merupakan representasi kata pada sebuah model ruang vektor multi dimensi. Setiap dimensi pada ruang berhubungan dengan sebuah kata. Posisi setiap axis di ruang ini tergantung pada frekuensi kata yang muncul pada teks. Kemiripan antara dua teks diukur berdasarkan jarak yang ada pada ruang vektor. Dimensi tersebut tidak harus linier, seperti perhitungan kata tunggal. Sebuah kata yang muncul sebanyak tiga kali, mungkin merupakan kata yang penting pada isi teks dibanding kemunculan kata yang hanya sekali. Tetapi kemunculan kata sebanyak tiga kali bukan berarti tiga kali lebih penting.

5. Similarity calculation

Pengukuran kemiripan (*similarity calculation*) merupakan perhitungan kemiripan antara dua teks yang dikerjakan pada model ruang vektor.

3. Hasil Dan Pembahasan

Berikut ini dipaparkan contoh dan cara penyelesaian yang dilakukan terhadap dua pendekatan yang dijelaskan sebelumnya. Contoh teks yang digunakan berasal dari contoh laporan. Contoh teks adalah sebagai berikut :

"Dump in file system when recording audio (see attached log)."

Penyelesaian dengan menggunakan pendekatan Ceska dan Fox sebagai berikut :

1. Tokenization

"Dump in file system when recording audio see attached log"

2. Stop-word removal (STR)

Menggunakan sebuah kamus *stop-word* yang berisikan daftar kata umum yang biasanya digunakan. Dalam contoh ini, kata yang umum adalah *in* dan *when*. Sedangkan kata *attached* dan *log* merupakan contoh kata yang sering muncul pada laporan tetapi tidak mempunyai arti penting.

"Dump file system recording audio"

3. Lemmatization (LM)

Pada contoh ini, bentuk dasar dari kata *recording* adalah *record*.

"Dump file system record audio"

4. Number Replacement (NMR)

Tidak terdapat angka yang akan dijadikan simbol dummy, sehingga bentuk teks sama seperti proses sebelumnya.

"Dump file system record audio"

5. Synonym Recognition (SYR)

Melakukan pengecekan di WordNet, apakah terdapat persamaan kata atau sinonim dari kalimat yang diperiksa.

6. Word Generalization (WG)

Melakukan proses penempatan kata-kata khusus menjadi kata-kata yang lebih umum.

Penyelesaian dengan menggunakan pendekatan Runeson, Alexandersson dan Nyholm sebagai berikut :

1. Tokenization

Seluruh tanda baca diabaikan.

"Dump in file system when recording audio see attached log"

2. Stemming

Seluruh imbuhan dihapus.

"Dump file system when record audio see attach log"

3. Stop-Words Removal

Hilangkan kata-kata yang secara umum digunakan atau kata-kata yang sering muncul tapi tidak mempunyai arti penting. Dalam contoh ini, kata *attached* dan *log*.

"Dump file system record audio"

4. Sinonim

Jika dokumen yang digunakan adalah dokumen besar yang dikirim oleh orang yang berbeda-beda, maka akan banyak terdapat kata yang berbeda digunakan untuk menyatakan sebuah arti kata. Untuk kebutuhan seperti ini maka perlu ditambahkan sebuah kamus. Kamus ini berisikan kata-kata yang dapat saling dipertukarkan atau dengan perkataan lain, kamus ini berisikan daftar

persamaan kata. Kamus ini diterapkan setelah melakukan *stop words removal*.

5. Pengukuran Kemiripan

Untuk mengukur kemiripan antara teks asli dengan teks yang dicurigai, dapat menggunakan model ruang vektor dengan menggunakan pengukuran *Cosine*, *Dice*, dan *Jaccard*.

Perbandingan pendekatan yang dilakukan Ceska dan Runeson

Dari uraian di atas, terlihat perbedaan dari kedua pendekatan yang ada. Perbedaan-perbedaan tersebut diantaranya :

1. Jumlah tahapan yang dilakukan dalam pra pemrosesan. Proses yang dilakukan oleh Ceska lebih banyak satu tahapan dibandingkan yang dilakukan oleh Runeson. Tahapan untuk melakukan pra pemrosesan teks yang dilakukan oleh Ceska sebanyak enam langkah. Sedangkan pra pemrosesan yang dilakukan oleh Runeson sebanyak lima langkah.
2. Teknik lain yang digunakan. Ceska menggunakan beberapa teknik tambahan sebagai alat bantu diantaranya menggunakan kamus *stop-word* pada saat melakukan tahapan *stop-word removal*. Kamus lemma digunakan pada saat melakukan tahapan *lemmatization*. Kamus *WordNet* yang digunakan pada saat pemeriksaan sinonim kata. Sedangkan Runeson hanya menggunakan kamus pada saat memeriksa sinonim, setelah melakukan *stop-word removal*.
3. Informasi dari hasil pra pemrosesan. Jumlah kata yang dihasilkan dari pra pemrosesan sampai sebelum pemeriksaan sinonim dengan pendekatan Ceska dan Runeson adalah sama yaitu sebanyak lima kata ("Dump file system record audio"). Tetapi tahapan yang dilakukan oleh Ceska lebih banyak satu langkah dibandingkan tahapan yang dilakukan oleh Runeson. Ceska membutuhkan empat langkah sedangkan Runeson membutuhkan tiga langkah untuk menghasilkan teks yang akan diperiksa sinonimnya.
4. Kekurangan. Kelemahan dari pendekatan yang dilakukan Ceska adalah adanya tahapan *number replacement*. Pendekatan ini mungkin sangat tepat untuk beberapa kasus. Misalnya mahasiswa mengirimkan tugas yang merupakan hasil plagiarisme yang berisikan analisis ekonomi suatu perusahaan. Hal tersebut sangat mudah untuk menggunakan hasil karya orang lain dengan menggantikan angka numerik dengan kombinasinya yang diperlukan. Tahapan *number replacement* bisa menyebabkan kurang akurat apabila berhubungan dalam hal tanggal. Kelemahan pendekatan Runeson yaitu terdapat tahapan *vector space representation* dan perhitungan kemiripan dimana tahapan tersebut dapat dilakukan sekaligus seperti pada tahapan *synonym recognition* yang terdapat pada pendekatan yang dilakukan oleh Ceska.

5. Kelebihan

Kelebihan dari pendekatan Ceska adalah penggunaan beberapa teknik tambahan seperti kamus stop-word, lemma dan wordnet yang sangat membantu pada beberapa tahapan pra pemrosesan. Kelebihan pendekatan yang dilakukan Runeson adalah sangat sederhana. Kamus sinonim diperlukan setelah melakukan tahapan *stop-word removal*.

4. Kesimpulan

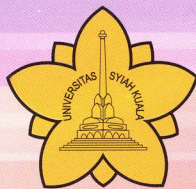
Tahapan pra pemrosesan dibutuhkan sebelum melakukan deteksi plagiarisme dengan menggunakan teknik *natural language processing*. Dari pendekatan pra pemrosesan teks yang dibahas maka pendekatan yang dilakukan oleh Cesko lebih baik dibandingkan pendekatan yang dilakukan oleh Runeson. Dengan penambahan teknik lain yaitu penggunaan kamus stop-word, kamus lemma dan wordnet sangat membantu dalam pra pemrosesan kata,

Penelitian berikutnya akan menguji apakah pendekatan pra pemrosesan yang dilakukan oleh Cesko berikut tahapan setelah pra pemrosesan dengan teknik *natural language processing* dapat diterapkan untuk dokumen berbahasa Indonesia.

REFERENSI

- [1] Aiken, A., 1994, *MOSS : A System For Detecting Software Plagiarism*, Stanford University. Diakses pada 1 Februari 2012 dari <http://theory.stanford.edu/~aiken/moss/>.
- [2] Androutsopoulos, I., and Malakasiotis, P., 2010, *A Survey of Paraphrasing And Textual Entailment Methods*, Journal Of Artificial Intelligence Research 38, 135-187.
- [3] Badge, J., and Scott, J., 2009, *Dealing With Plagiarism In The Digital Age*, University Of Leicester, pp. 1-18.
- [4] Ceska, Z., and Fox, C., 2009, *The Influence Of Text Pre-processing On Plagiarism Detection*, International Conference RANLP 2009 - Borovets, Bulgaria, pp. 55-59.
- [5] Chong, M., Specia, L., and Mitkov, R., 2010, *Using Natural Language Processing for Automatic Detection Of Plagiarism*, Proceedings Of The 4th International Plagiarism Conference (IPC-2010), Newcastle-upon-Tyne, UK.
- [6] Clough, P., 2003, *Old And New Challenges In Automatic Plagiarism Detection*, National Plagiarism Advisory Service, February, pp. 391-407.
- [7] Leung, Chi-Hong and Yuen Yan Can., 2007, *A Natural Language Processing Approach To Automatic Plagiarism Detection*, Proceeding Of The 8th ACM SIGinformation Conference On Information Technology Education SIGITE 07.
- [8] Lukashenko, R., Graudina, V., and Grundspenkis, J., 2007, *Computer-Based Plagiarism Detection Methods And Tools : An Overview*, International Conference On Computer Systems And Technologies - *CompSysTech'07*.
- [9] Runeson, P., Alexandersson, M., and Nyholm, O., 2007, *Detection Of Duplicate Defect Reports Using Natural Language Processing*, 29th International Conference On Software Engineering, ICSE'07, pp. 499-510.

Organizer :



JURUSAN TEKNIK ELEKTRO
UNIVERSITAS SYIAH KUALA



DEPARTEMEN TEKNIK ELEKTRO
UNIVERSITAS SUMATERA UTARA



9 772088 998005